DIRECTIONS OF MATHEMATICAL RESEARCH IN NONLINEAR CIRCUIT THEORY

Richard Bellman
Mathematics Division
The RAND Corporation

P-1802

23 September 1959

SUMMARY

In what follows we wish to present a potpourri of problems of some interest and difficulty arising in the field of nonlinear circuit theory. Perhaps the only sensible way to catalogue scientific problems is in terms of "solved or unsolved." Yet this classification is itself a very subjective one, dependent upon the times and the fashions. Recall the dictum of Poincaré that the solutions of one generation are the problems of the next.

In what follows, we have for the sake of convenience attempted to group categories of problems under the headings of "descriptive," "control," "stochastic," and so forth. Convenient as some of this nomenclature is, it should be regarded with a certain amount of suspicion. Most significant problems blithely cut across these artificial boundaries within fields of specialization, and within science itself.

In these days of rapidly and dramatically changing technology, it would be rather brash to attempt to predict the type of mathematics that will be most urgently required even ten years from now. It is, however, fairly safe to look about and note the requirements of the present and of five years back. The difficulties that abound render a certain time lag inevitable, and it may well be that new scientific developments may render fields obsolete and matnematical solutions for problems within those fields unnecessary before they are even obtained.

CONTENTS

DIRECTIONS OF MATHEMATICAL RESEARCH IN NONLINEAR CIRCUIT THEORY

Richard Bellman

INTRODUCTION

In what follows we wish to present a potpourri of problems of some interest and difficulty arising in the field of nonlinear circuit theory. Perhaps the only sensible way to catalogue scientific problems is in terms of "solved or unsolved." Yet this classification is itself a very subjective one, dependent upon the times and the fashions. Recall the dictum of Poincaré that the solutions of one generation are the problems of the next.

In what follows, we have for the sake of convenience attempted to group categories of problems under the headings of "descriptive," "control," "stochastic," and so forth. Convenient as some of this nomenclature is, it should be regarded with a certain amount of suspicion. Most significant problems blithely cut across these artificial boundaries within fields of specialization, and within science itself.

In these days of rapidly and dramatically changing technology, it would be rather brash to attempt to predict the type of mathematics that will be most urgently required even ten years from now. It is, however, fairly safe to look about and note the requirements of the present and of five years back. The difficulties that abound render a certain time lag inevitable, and it may well be that new scientific developments

may render fields obsolete and mathematical solutions for problems within those fields unnecessary before they are even obtained.

We have made no attempt to present all of the fields of mathematical research that stretch ahead so invitingly. Rather, we have fastened our attention upon those that seem most intriguing or most significant in our own eyes, and those in which we have meandered to some extent or other.

Finally, we make no pretense of knowing how to tackle most of the major problems we pose. If anything, we may conclude that new approaches are vitally needed at the present time to tackle the truly formidable problems we face. There is, in consequence, an enormous opportunity for the original mind to made fundamental contributions.

DESCRIPTIVE PROCESSES

1. Descriptive Processes

We shall consider two essentially different types of processes. One we shall call descriptive and the other control. In the first case, we shall consider a physical system whose state is described by a functional equation of the form

$$\frac{dx}{dt} = F(x), \quad x(0) = c. \tag{1}$$

Here x is an N-dimensional vector and $F(x)$ is a functional of $x(t)$. The problem is then that of determining the behavior of $x(t)$ over all time. This field contains the classical theory of differential equations, and, as we shall see, much more.

Subsequently, we shall define what we mean by a control process in more precise fashion. Meanwhile, let us think of a process of this nature as one in which we alter the structure of the system in various ways in order to obtain a more desirable behavior of the system over time.

Within the category of descriptive processes, we shall make further splittings into deterministic and stochastic, linear and nonlinear, and finally, low- and high-dimensional.

It is not hard to believe that we could devote a paper of

this size and more to any one of these subdivisions. Consequently, if we appear to skim and dart or to stroll in seven league boots, we hope that the reader will forgive us.

2. Differential Equations

Let us begin with the traditional subject of differential equations. Let x be an N-dimensional vector determined by the vector differential equation

$$(1) \qquad \frac{dx}{dt} = g(x), \quad x(0) = c.$$

The basic challenge is that of predicting the behavior of $x(t)$ as t increases. This is the classical descriptive process.

For some reasons which we shall enter into below this is a never-ending challenge which in all probability will never be completely answered. However, for reasons which we shall also describe below, it may be a challenge to which we may not wish to respond.

In the following sections, we shall discuss linear and nonlinear differential equations, analytic and computational aspects, deterministic and stochastic versions, and, finally, some aspects of the curse of dimensionality.

3. Linear Differential Equations—Analytic Aspects

The behavior of the solution of a differential equation is quite readily determined when we possess some explicit representation such as $\sin(\omega t + \phi)$ or $J_0(t)$, or even a contour integral such as

(1) $$u(t) = \int_C e^{zt-z^4}\,dz.$$

In general, differential equations introduce new functions which cannot be represented in terms of the familiar transcendents of Whittacker and Watson.

As simple an equation as the second order linear differential equation

(2) $$a'' + a(t)u = 0,$$

cannot be integrated in terms of quadratures and the usual functions of mathematical physics when $a(t)$ is a general function; cf. [1]. Furthermore, even if $a(t)$ has a quite simple form, there is no guarantee that the solutions will possess any simple structure. Thus the equation

(3) $$u'' + (a + b\cos t)u = 0,$$

the Mathieu equation, presents many intricacies of analysis, and essentially has a theory of its own.

It follows that new techniques must be devised to enable us to derive information concerning the nature of the solution directly from the properties of the coefficients, without the intervention of explicit analytic solutions. An enormous amount of effort has been expended in this activity, stimulated by the classical Sturm–Liouville problems of mathematical physics, and by the newer Schrodinger equation; see [2], [3], for many references.

The study of linear systems is of particular importance in conjunction with stability analysis, [2]. Leaving aside the case of constant coefficients, by no means a trivial case as we shall see below, the next two most important types of linear equations to consider are those with periodic coefficients and those with almost-periodic coefficients.

Let us discuss these equations quite briefly. For the case of periodic coefficients, the result of Floquet yields a foothold which can occasionally be further exploited. Given the equation

$$\frac{dx}{dt} = A(t)x, \tag{4}$$

where $A(t)$ is periodic, say of period 1, the aforementioned result asserts that every solution of (4) has the form

$$x = e^{Bt}p(t), \tag{5}$$

where $p(t)$ is a periodic vector with period 1 and B is a constant matrix. This is an existence theorem which of itself furnishes no information as to the nature of B.

Any results beyond this are come by hard. For an extensive survey of results and techniques, see [4]. Let us make one remark in this connection. Suppose that $A(t)$ is not only singly-periodic in the real variable t, but doubly-periodic as a function of the complex variable t. In this case, one can theoretically obtain an explicit representation for x in terms of doubly-periodic functions. This result of Hermite, [5], has not been exploited. It seems reasonable to

suppose that a good deal of information concerning the Mathieu functions and related functions can be obtained using the fact that the Jacobian elliptic functions reduce to the usual sine and cosine as the modulus approaches zero.

For the case of almost-periodic coefficients, the situation is very discouraging, despite the determined endeavors of a number of mathematicians. A little known paper by Shtokalo, [6], contains some interesting results which seem capable of extension, but apart from this there are few leads. It would seem that the class of almost-periodic functions is too wide a class, and that one must restrict the coefficients to some class intermediate between this and the tamer class of periodic functions.

4. Nonlinear Differential Equations

Whatever the difficulties are in the linear case, they fade before the frightening aspects of nonlinearity. When treating a linear equation, we enter the lists armed with the fact that the manifold of solutions is a linear space of finite dimension, the dimension of the system. When dealing with nonlinear equations, we possess no such advantage. In the absence of additivity, the superposition principle, the standard methods of analysis become inoperative.

New techniques must be employed. One is the idea of a regular solution, a solution finite for $t \geq t_0$. This very simple and fruitful idea, introduced by Borel, and developed

by Lindelof, Hardy, and Fowler (see [2] for results and references), enables us to study in great detail the behavior of important classes of solutions of polynomial equations of the form

(1) $$g(u,u',t) = 0,$$

and

(2) $$u'' = h(u,u',t).$$

In particular, equations of the Emden–Fowler type,

(3) $$u'' \pm u^a t^b = 0,$$

can be intensively analysed.

Another is the topological approach of Poincaré which we shall discuss below.

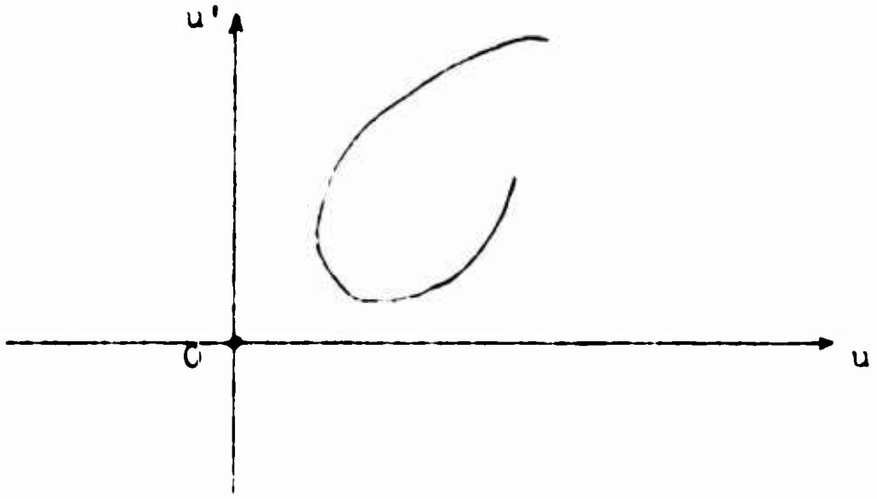
5. Nonlinear Differential Equations—Geometric Aspects

To Poincaré is due the elegant idea of using fundamental topological properties of curves and vector fields to study the properties of solutions of nonlinear differential equations. An excellent exposition of these ideas is given in [7].

Consider the equation

(1) $$u'' = g(u,u'),$$

and the graph of a solution in the phase-plane, the (u,u')-plane.



As pointed out by Poincaré, the particular solutions which enable us to sift and sort the diverse kinds of solutions of (1) are the periodic solutions, which in the phase plane represent closed curves. Assuming, as is true in the most interesting cases, that solutions of (1) are uniquely specified by values of u and u' at some time t_0, we see that the graphs of solutions in the phase plane cannot intersect each other.

Upon this simple fact hinges the success of the topological analysis of the solutions of (1). As soon as we turn to the study of fourth order systems

$$u^{(4)} = g(u,u',u^{(2)},u^{(3)}), \tag{2}$$

or to the study of two coupled equations

$$u^{(2)} = g(u,u',v,v'), \tag{3}$$

$$v^{(2)} = h(u,u',v,v'),$$

a more natural situation, the situation changes radically.

In four-dimensional phase space, solutions can loop, intertwine and knot in the most intricate and abandoned fashion. Such is the degree of complexity encountered in the study of general curves and surfaces in four-dimensional space that far from expecting topological results in this area to contribute to the study of the particular class of curves and surfaces generated by differential equations, one should expect the contributions to flow in the other direction.

6. Discussion

Little is known about the behavior of solutions of systems of the form of (6.3), and the prospects are not encouraging for the immediate future. Despite intensive research over the last twenty years, no breakthroughs of any significance have been made.

Sometime in the future, as has often happened in the past, a theory started from a quite different origin may reach a summit from which we shall be able to look down and perceive a number of paths to peaks which are from our view buried in clouds. More likely, we may never reach such a vantage point. This lack of mathematical ability may affect our technological development to the degree that other techniques based upon different mathematical models will be developed. The problems we have mentioned concerning linear and nonlinear differential equations may ultimately be relegated to the mathematical limbo that contains so many problems of live interest in the eighteenth

and nineteenth centuries.

If the present deadlock continues, we will be forced to the reluctant conclusion that the classical vein of differential equations is played out. Fortunately, the modern theory of control processes provides a natural continuation with unlimited regions for research at the moment.

This is another example of the by now fairly widely recognized fact that pure mathematics without the constant stimulation and enlargement by the physical world faces the dreer prospects of sterility and decadence.

7. Nonlinear Differential Equations--Computational Aspects

The inadequacy of our analytic techniques in the face of nonlinearity, and perhaps even the non-existence of such techniques, compels us to focus the greater part of our attention upon the computational solution of differential equations.

Here, we face the problems of feasibility, accuracy and stability. The overall question is that of adapting an analytic algorithm such as successive approximations to the task of supplying numerical answers to numerical questions.

An enormous amount of research, and mathematical experimentation, is required in this area. It must be realized that significant computational advances can only be consequences of significant analytic advances. The two go hand-in-hand, with one acting as an incentive to the other.

Let us discuss briefly two of the new techniques developed

over the last few years. One is the concept of quasilinearization, [8], [9], and the other is that of sequential computation, [10].

To indicate the basic idea of quasilinearization, let us consider the problem of solving the two point boundary-value problem

$$(1) \qquad u'' + g(u) = 0,$$

$$u(0) = u(a) = 0.$$

Let us suppose that a is small enough so that there is a unique solution, and that $g(u)$ possesses a second derivative.

Let u_0 be an initial approximation, and let the sequence $\{u_n\}$, $n = 1,2,\ldots,$ be generated by means of the equation

$$(2) \qquad u_n'' + g(u_{n-1}) + (u_n - u_{n-1})g'(u_{n-1}) = 0,$$

$$u_n(0) = u_n(a) = 0.$$

It can be shown that with this method of successive approximations that the convergence is quadratic, i.e., $|u_n - u_{n-1}| = O(\delta^{2^n})$, rather than geometric, $|u_n - u_{n-1}| = O(\delta^n)$, and, in some cases, that it is also monotonic. For details, see [9].

This method has also been applied successfully to different types of partial differential equations; see [9], [11].

The concept of sequential computation is related to the existing theory of computation in the very same way that

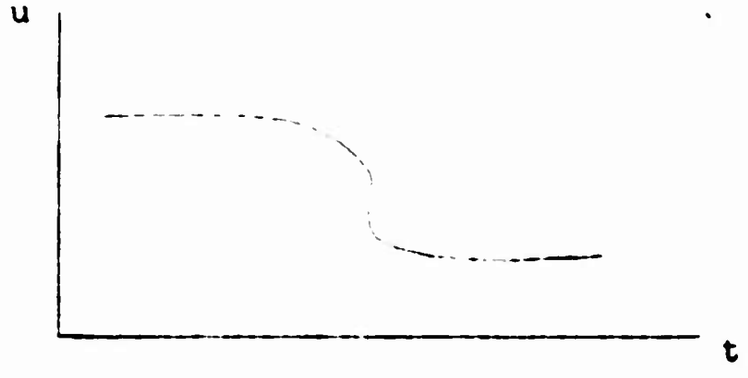
sequential analysis compares to statistical theory prior to Wald. At the present time, when considering an equation such as

$$(3) \qquad \frac{du}{dt} = g(u), \quad u(0) = c,$$

we generally use a discrete version such as

$$(4) \qquad \frac{u_{n+1} - u_{n-1}}{2\Delta} = g(u_n), \quad u_0 = c,$$

where $u_n = u(n\Delta)$. If, over certain portions of the t-interval, the function $u(t)$ exhibits too wild a behavior, e.g.,



we decrease the grid size over this interval. This is usually done in an a posteriori fashion. By this we mean that first we run through the solution with a grid size Δ, and then begin to worry if we observe a behavior such as that shown above.

What would be much more efficient would be to have a computational technique which automatically adapts itself to the behavior of the solution that is being computed. One way of doing this would be to make Δ dependent upon the slope and

the curvature of the solution. In place of a recurrence relation such as (4), we would have two relations such as

$$(5) \qquad \frac{u_{n+1} - u_{n-1}}{\Delta_n} = g(u_n), \quad u_0 = c,$$

$$\Delta_n = h(u_n, u_{n-1}, u_{n-2}).$$

For some further details, see [10].

This is merely one aspect of the general concept of adaptive algorithms rather than fixed algorithms. The area is almost completely unexplored.

8. Linear Differential Equations—Dimensionality Difficulties

Leaving for the moment the rocky region of nonlinearity, let us discuss a major problem associated with linear differential equations.

Suppose that we have reduced the problem with which we are concerned to that of determining the solution of the linear differential equation

$$(1) \qquad \frac{dx}{dt} = Ax, \quad x(0) = c,$$

where A is a constant matrix. If the dimension of x is moderate, say ten or twenty, we can consider the solution to be routine. If the dimension is large, say several hundred or a thousand, then we have dimensionality difficulties.

There are now several different techniques that we can employ to replace a direct method of solution based upon the explicit analytic representation.

In the first place, we can examine the original problem and see whether or not it can be reformulated in such a way as to lend itself to a computational solution in terms of vectors of considerably lower dimension. In a number of cases, this can be done, see [12], [13], [14]. Now that we are seriously beginning to consider systems of quite high dimension, it is essential that the mathematical models employed in the past for systems of low dimension be systematically examined from the standpoint of computational and analytic convenience. We shall return to this matter below.

The basic idea is that of reducing a process of high dimension by a sequence of processes of low dimension. This is the basic idea of dynamic programming, [15], and is the essence of a new method due to Kron, the method of "tearing," which seems to have very interesting possibilities; see Roth, [16].

Along similar lines are the relaxation methods of Southwell, and the "flooding method" of Boldyreff, [17]. Much remains to be done to make these techniques generally applicable and efficient. One has the feeling that mathematicians have really understood the nature of the processes with which we deal on the whole so clumsily, the analytic and computational algorithms will consist of very simple operations, the proof of whose efficacy will depend upon quite sophisticated concepts. At the present time, we largely depend upon very complex operations resting upon very rudimentary concepts.

9. The Inverse Problem

So far we have assumed that the equation is given and that it is required to determine the behavior of the solution over time. In many situations, the reverse is the case. Given the behavior over time of a vector $x(t)$, it is required to determine the possible equations which can produce this behavior.

This problem has been intensively investigated in connection with Sturm–Liouville equations, (see [18], [19]) and is closely connected with the "black box" problem we shall discuss subsequently.

10. Time-lags and Hereditary Effects

We meet all the foregoing problems and some additional ones if we take into account delay, time-lags, retardations, and hereditary effects in general.

The simplest equations arising in this way are the differential-difference equations, which have the form

$$\frac{dx}{dt} = g(x(t), x(t - \tau_1), \ldots, x(t - \tau_m)). \tag{1}$$

The appropriate initial condition is now an interval condition

$$x(t) = h(t), \quad 0 \leq t \leq \tau_m. \tag{2}$$

In many cases, more complicated equations are necessary to describe the state of the system, equations such as

$$(3) \qquad \frac{dx}{dt} = g(x(t), \int_{-\infty}^{t} k_1(t,s)x(s)ds, \ldots, \int_{-\infty}^{t} k_m(t,s)x(s)ds).$$

The most interesting of these are the renewal equations

$$(4) \qquad x(t) = g(t) + \int_0^t k(t-s)x(s)ds,$$

arising in the theory of queues, counters, in the theory of branching processes, and in prediction theory.

For a discussion of equations of this nature, see [20], [21].

11. Stochastic Processes

It should be recognized initially that the formulation of a physical process in deterministic or stochastic terms is a matter of mathematical convenience rather than a consequence of the intrinsic structure of the process. Let us, however, bypass the really deep problems connected with the advantages or disadvantages of one formulation or another, and instead indicate some of the many new and challenging problems that arise once we have agreed to describe the behavior of a circuit in terms of differential equations with stochastic elements.

We shall consider only ordinary differential equations, although the corresponding problems for partial differential equations are of great importance in other parts of mathematical physics, e.g., turbulence theory.

Let us consider four particular problems. It would be as easy to present forty problems, since the field is practically unexplored. A certain amount of effort has been devoted to

the study of the linear equation

$$(1) \qquad L(u) = r(t),$$

where L is a deterministic operator and $r(t)$ is a random function. The results that are obtained and techniques that are employed are rather straightforward, and thus not particularly challenging. Far more interesting is the situation in which the operator itself has a stochastic structure. Questions of this type arise in the consideration of circuits with unreliable elements, or in the study of circuits subject to external influences of incompletely known type which affect the structure of the circuit. For an interesting discussion of results obtained to date for the case of linear differential operators, see [22], where many further references will be found.

In order to illustrate some of the difficulties in the study of this problem, let us consider perhaps the simplest version, that of a linear difference operator.

The equation we wish to study is

$$(2) \qquad x(n+1) = Z(n)x(n),$$

where $x(n)$ is an N-dimensional vector and $Z(n)$ is an $N \times N$ random matrix. It is sufficient to consider the simplest case where the $Z(n)$ are independent random matrices.

It is not difficult to obtain analytic expressions for the moments of the components of $x(n)$, using the technique of

Kronecker products of matrices, see [23], [24], [25]. The fact that Kronecker products of matrices bypass the difficulties induced by noncommutativity of matrices has not been sufficiently exploited in the study of linear stochastic systems of this nature.

Furthermore, there is little difficulty in extending these results to the case where the $Z(n)$ are correlated in some Markovian fashion, and in passing to the limit and obtaining corresponding results for linear differential equations, although no results of this type have been published.

What is really difficult is a more precise determination of the asymptotic behavior. We expect the solutions to exhibit exponential behavior, and the moments that are obtained illustrate this. It follows that it is the moments of the logarithm of the components that yield more detailed information. For example, we would like to be able to find the asymptotic behavior of

$$(3) \qquad \mathrm{Exp}(\log x_1(n))^K$$

as $n \to \infty$. Here $x_1(n)$ is the first component of $x(n)$. Although some small steps in this direction have been taken, see [25], no satisfactory methods or results exist at the moment.

For our second problem, let us turn to the comparison of discrete and continuous versions of stochastic processes. Parenthetically, let us observe that it is often felt that the discrete version of a process is an approximation introduced

either for computational purposes, as when we replace differential equations by difference equations, or for conceptual simplicity. It is worth emphasizing the fact that the reverse may well be true, and certainly is true in a number of control processes. The actual process will often be discrete, but a continuous process will be used sometimes for analytic convenience, but more often by very force of tradition. If one takes into account the fact that the solution will ultimately be obtained by means of a digital computer, then there may be a loss of efficiency, and even error, introduced by conversion of a discrete process into a continuous process for the sake of analysis, and then by reconversion to a discrete process for the sake of computation. The fact that the original discrete process and the final discrete process may well be different processes can often cause grief. On the other hand, this may be an advantage. We do not wish to make any rigid statement concerning the advisability or inadvisability of one procedure or another. What we do wish to emphasize is that it is essential to be aware of these matters, so that one can follow the procedure which is most advantageous in any particular process.

It is particularly important to be cognizant of these facts at the present time since the digital computer has introduced a flexibility into formulation and solution of mathematical problems which was certainly not present even as recently as ten years ago.

Since we know that the same physical process can be formulated in many different fashions, discrete or continuous, deterministic or stochastic, and so on, it is essential that we know the relations between these different formulations. In particular, it is essential that different formulations yield the same numbers. Perhaps even more important would be a knowledge of when they do not, since this would be a most valuable clue concerning the true nature of the process.

For the case of deterministic processes, a great deal is known concerning the relation between discrete and continuous versions. For example, given the differential equation

$$\frac{dx}{dt} = g(x), \quad x(0) = c, \tag{4}$$

and the discrete approximation

$$x_{n+1} - x_n = g(x_n)\Delta, \quad x_0 = c, \tag{5}$$

where $x_n = x(n\Delta)$, quite simple conditions imposed upon $g(x)$ guarantee that $x(n\Delta) \to x(t)$ as $\Delta \to 0$, $n\Delta \to t$. Essentially, these are the same conditions that ensure existence and uniqueness of solution of (4), although many refinements exist. For specific results and further references, see [2], [26].

For the case of stochastic processes, this problem seems not to have been studied at all. A question of some interest is the following: What is the relation between the solutions of the two equations

(6) $$x'(t) = g(x(t),r(t)), \quad x(0) = c,$$

and

(7) $$x_{n+1} - x_n = g(x_n,r_n)\Delta, \quad x_0 = c,$$

($x_n = x(n\Delta)$, $r_n = r(n\Delta)$), where $r(t)$ is a random function with a given distribution? If $r(t)$ is uniformly bounded, there is no difficulty in applying the usual methods. The interesting case is that where $r(t)$ is unbounded.

Let us turn back to the "black box" problem. Given a linear operator, L, such as a linear differential operator, we can characterize it in simple fashion by examining the solutions of the inhomogeneous equation

(8) $$L(u) = e^{i\omega t}$$

as ω ranges over a set of values. What does one do for a nonlinear differential operator? This is a problem which has been studied to some extent, see [2], [28].

Is it true that nonlinear operators are more appropriately studied by means of stochastic forcing terms than by means of deterministic ones? For example, should one use the equation

(9) $$N(u) = r(t)$$

where $r(t)$ is a random function with an appropriate distribution, rather than an equation of the form

(10) $$N(u) = e^{i\omega t},$$

or an equation with another type of deterministic forcing term?

Closely related to this problem is the final problem we wish to pose. The importance of the Gaussian distribution lies in its relative invariance under a linear transformation. By this we mean that if x and y are both Gaussian, then $x + y$ is Gaussian. Another way of looking at this is that the linear transformation

$$T(x) = x + r \tag{11}$$

is a particularly simple one to study when r is Gaussian.

Viewed in this way, the natural question to ask is: What class of distributions should be used when we study more general nonlinear transformations of the form

$$T(x) = g(x,r) \tag{12}$$

Furthermore, what type of limiting distributions should be expected when the transformation is iterated repeatedly?

These questions seem of great difficulty. It may well be that the way to approach them is to do some mathematical experimentation with digital computers, using particular types of simple transformations and random variables. Work of this type has been carried out by Stein and Ulam at Los Alamos, but not published as yet.

CONTROL PROCESSES

1. Control Processes

Let us now leave the realm of descriptive processes and turn to the study of control processes. There is little point at the present time, in view of the many different uses of these words in many diverse fields, in attempting to make precise what is meant by the term "control process." Let us simply and intuitively say that we wish to use it to describe any physical process in which the original behavior of the system is influenced in some way in order to force it to operate in a more desirable fashion.

A priori, it might be imagined that the study of control processes is more complex than that of descriptive processes. To some extent, this is true. On the other hand, the additional information furnished by the requirement that we exert control in a most efficient fashion often yields a considerable simplification. It is generally the case that optimal policies are simpler than obvious feasible policies, once their true nature is understood.

It may very well be the case that our excursion into this more difficult field will ultimately yield the vantage points required to resolve some of the truly obdurate problems in the theory of descriptive processes.

In any case, the theory of control processes is replete with fascinating, challenging mathematical problems of the utmost significance to contemporary society. What more can a mathematician ask?

2. Variational Analysis

Once again beginning in a traditional way, let us consider the following analytic version of a feedback control process. Consider a physical system S whose state at time t is described by the vector $x(t)$ which satisfies a differential equation

$$(1) \qquad \frac{dx}{dt} = g(x,y), \quad x(0) = c.$$

Here $y = y(t)$ is the control vector which is to be chosen so as to minimize the functional

$$(2) \qquad J(x,y) = \int_0^T h(x,y)dG(t).$$

If $dG(t) = g(t)dt$, we have a variational problem of the usual type. If $G(t)$ is a step-function with a single jump at T, we have the problem of minimizing

$$(3) \qquad J(x,y) = h(x(T),y(T)),$$

a terminal control problem. In many cases, we have a combination of both types.

When these problems are approached by the classical methods of the calculus of variations, they lead to nonlinear differential equations of dimension double the dimension of x, with two-point boundary conditions. These problems are analytically intractable and require a great deal of effort and good fortune if a reliable numerical result is desired.

A new approach to questions of this nature along the lines

of dynamic programming, [10], [13], permits some advances to be made. However, for problems of even moderate dimension any routine application of these techniques also runs into acute computational difficulties.

We can consider the problem of obtaining feasible algorithms for the computational solution of realistic feedback control processes to be one of the challenging and important fields of research in the immediate future. Whatever the methods in use now, better ones are needed.

We shall have more to say on the subject below.

3. Constraints

To turn the screw a little tighter, let us point out that in many significant situations, we do not have complete freedom in a choice of y. The simplest and most natural type of constraint is one of the form

$$|y_i| \leq m_i, \quad i = 1,2,\ldots,N, \quad 0 \leq t \leq T. \tag{1}$$

There is now considerable difficulty in applying variational analysis, and even the most simply posed analytic problems possess obstacles in the path of their solution; see [10], [29].

It would be extremely valuable to have available the complete solution to certain prototype problems. For example, consider the problem of minimizing the functional

$$J(x,y) = \int_0^T \left[(x,Bx) \cdot (y,Cy)\right] dG(t) \tag{2}$$

over all y where

(3) $$\frac{dx}{dt} = Ax + y, \quad x(0) = c,$$

and y is subject to (1). Can this problem be completely resolved?

4. Bang-bang Control

A very important type of problem that arises is many areas is that of forcing a system back to an equilibrium position in a minimum time. A great deal of effort has been devoted to the study of linear systems, such as that of (3.3), with constraints of the form of (3.1), and the problem has been fairly well tamed; see [30], [31], where further references may be found.

For nonlinear systems of small dimension, the functional equation technique of dynamic programming may be employed, see [31]. For nonlinear systems of moderate or high dimension, no techniques exist for either analytic or computational solution at present.

Problems of this type are actually only particular cases of variational problems of implicit nature. These are characterized by the property that the form of the criterion function itself depends upon the policy that is pursued.

For example, in place of a fixed functional of the form

(1) $$J(y) = \int_0^T h(x,y)dt,$$

we may envisage one of the form

$$(2) \qquad J(y) = \int_0^{T(x,y)} h(x,y)\,dt.$$

"Bang-bang" control corresponds to $h(x,y) \equiv 1$, which means that $J(y)$ measures the duration of the process.

5. Successive Approximation

The standard analytic technique for circumventing and overcoming the usual intractable aspects of the equations of analysis is the method of successive approximations. Although this method has been applied with perseverance and success to the study of descriptive processes, for some reason or another little has been done with it in the study of control processes.

From the very nature of the feedback control process, one would suppose that successive approximations would be ideally suited to the study of these processes.

Let us then base our method of successive approximations around this fact. Starting with some initial approximation to a policy, y^0, and an approximation to the state vector, x^0, let us replace $J(y)$ by its expansion around these functions x^0 and y^0. In this way, retaining only terms up to and including the second degree, we obtain a new functional which is quadratic in x and y. Similarly, let us expand the function $g(x,y)$ around the point x^0, y^0, retaining only terms up to the first degree. In this way the describing equations are linearized.

Solving this approximate variational problem in analytic

terms, we obtain a new policy vector, y^1, and a new state vector x^1. This process can now be continued. We now face the usual problems concerning convergence, rate of convergence, convergence to a relative instead of absolute extremum, and so on.

There are a number of other ways in which the concept of successive approximations can be applied; see [9], [10], [32], [15]. A quite different approach will be described below.

6. Approximation to Functions

Let us continue on this theme of dimensionality. We have been forced to think of various types of approximations because of the impossibility of dealing with functions of a large number of dimensions in any direct fashion. Because of the limited fast memory of present-day digital computers, and of any that we can contemplate in the immediate future, we must renounce any routine approach based upon the tabulation of functions of say ten variables over the gridpoints of a ten-dimensional region.

Instead, if we are dealing with smooth functions, we can contemplate a more analytic description in terms of polynomials or in terms of orthonormal expansions. These latter are preferable since they usually possess superior convergence properties.

If we use an approximation of the form

$$(1) \qquad u(s,t) = \sum_{m,n=0}^{N} a_{m,n} v_m(s) v_n(t)$$

over the unit square, we require $N(N+1)/2$ coefficients to determine the function $u(s,t)$. If we tabulate the values of the function at the gridpoints $(k\Delta, \ell\Delta)$, we require $1/\Delta^2$ values. As we increase the dimension of the function u, the number of gridpoints goes up exponentially, while the number of coefficients required to determine the function to the same degree of accuracy by means of an approximation of the form in (1) goes up at a much slower rate--provided, of course, that the function is smooth. In many important applications, this will be the case.

It is important to note that this type of approximation can be used not only in the study of control processes, but in the study of descriptive processes as well. A brief discussion of a particular control process treated in this fashion will be found in [33].

7. Stochastic Control

As soon as we turn to the subject of stochastic control processes, we face the problem of treating variational questions involving stochastic functionals and stochastic differential equations.

The functional equation technique of dynamic programming, [10], [34], can be readily applied to treat these new types of analytic problems. As a matter of fact, the treatment is abstractly similar to that of deterministic control processes.

There are, however, a large number of mathematical questions in this field which have neither been precisely

asked nor precisely answered. Generally, we want a rigorous formulation of the problem of minimizing the functional

$$(1) \qquad J(y) = \operatorname*{Exp}_r \left[\int_0^T g(x,y,r)dt \right]$$

where x and y are connected by means of a stochastic differential equation

$$(2) \qquad \frac{dx}{dt} = h(x,y,r), \quad x(0) = c.$$

Here $\operatorname*{Exp}_r$ represents the expected value taken with regard to the random elements.

As usual, it would be worthwhile to begin with to consider the case of quadratic functionals subject to linear describing equations, since, as mentioned above, these can be used as the point of departure for successive approximations.

It would also be useful to treat the "bang-bang" control problem under the assumption that the describing equation is linear,

$$(3) \qquad \frac{dx}{dt} = Ax + y + r, \quad x(0) = c,$$

and we wish to minimize the expected time to get to the origin.

ADAPTIVE PROCESSES

1. Preliminaries

We have briefly discussed descriptive and control aspects of deterministic and stochastic processes. In the main, the mathematical questions that we have asked have been easy to pose, but difficult to answer. This is rather typical of classical analysis. In the next two parts of the paper, we wish to depart from the well-charted roads and follow some brambly trails. Our aim is to discuss some parts of nonlinear circuit analysis which give rise not only to questions which are difficult to answer, but difficult even to ask. As a matter of fact, to know what to ask and how to ask it is to be far along the path to a solution.

Basically, we wish to consider classes of problems which are not well-defined according to conventional tenets. One way of formulating problems of this nature precisely is furnished by the theory of games, [35], [36], either single-stage or multi-stage, [15]. We shall utilize the multi-stage aspect to provide another formulation. It must be emphasized that there is no definitive way of handling these new types of problems at the present time, and it is extremely unlikely that there ever will be.

The processes we wish to discuss are those in which we lack certain essential information which was given to us in our study of deterministic and stochastic processes. This information can be obtained, bit by bit, as the process un-

folds. Not only must we make decisions on the basis of the information we possess, but we must make decisions as to what information we wish to possess. Processes of this nature, we call "adaptive processes." Occasionally, the term "learning process" is employed.

In some of these processes, it may be senseless to speak about "optimal policies," but quite sensible to talk about feasible or efficient policies; cf., for example, [37].

2. Unknown Distributions

Rather than continue in these vague, foreboding terms, let us consider a stochastic control process governed by the equation

$$x_{n+1} = g(x_n, y_n, r_n), \quad x_0 = c, \tag{1}$$

where $\{r_n\}$ is a sequence of random variables with an unknown distribution.

What do we mean by an optimal control policy in a situation like this? One way out is to employ the theory of games, under the assumption that some malevalent influence is choosing the unknown distribution. This is unduly pessimistic, and, in any case, eliminates the possibility of our learning about the distribution of r_n as the process continues.

The question that faces us now is that of formulating this learning process in mathematical terms. We can proceed in various ways:

(2) (a) We can record the past history of the process, $\{x_k, r_k\}$, $k = 1, \ldots, -M$, and use this data to predict the distribution of r_0, and so on.

(b) We can use an a priori distribution for r_0, $dG(r)$, and use a fixed rule to modify $dG(r)$ on the basis of observation of r_0 or x_1.

(c) We can use an a priori distribution $dG(r,a)$, dependent on a finite set of parameters, $\{a\}$, and use a fixed rule to modify a on the basis of observation of r_0 or x_1,

and so on, and so on.

Not only do we face the task of justifying a choice of one model or another, but granted any particular formulation, we must endure formidable mathematical difficulties in our attempts to use these mathematical models to obtain numerical answers to numerical problems. In some cases, the functional equation technique of dynamic programming can be utilized; see [10], [38], [39], [40]. For an interesting application to prediction theory, see [41].

3. Turning the Screw Again

Whatever the difficulties raised by the foregoing, we can compound them by supposing

(1) (a) that the form of g in (2.1) is not known.

(b) that neither the form of g nor the dimension of x_n is known.

(c) that the criterion for determining an optimal policy is not precisely known.

(d) that the duration of the process is not known.

The domain of the new theory of adaptive processes consists of problems of this nature, and contains some other troglodytes which we shall dig out below. Some of these questions are discussed in [10].

It seems reasonable to expect that these totally new types of problems will require new methods and new concepts. One might surmise that the emphasis will be more upon policies and less upon functions, and more upon approximation in policy space, than in manipulation with functions and functionals.

THE CURSE OF DIMENSIONALITY

1. Introduction

Examining the various problems that we have skimmed over, lamenting all the while, it is clear that the conceptual, analytic and computational features become enormously more complex whenever the dimension of the system S becomes large. This is an example of the metaphysical axiom that a significant difference in quantity constitutes a difference in quality.

In this concluding section, we wish to touch lightly upon some new techniques that have been proposed in recent years to treat various problems involving large numbers of state variables, vast accumulations of data, and lack of basic information.

To our mind, this field is the most challenging of all the new areas that have opened up over the last ten years. It is well suited for the young research scientist since it requires new, original and daring ideas with no apparent payoff for erudition or experience. If anything, saturation with old ideas is a handicap.

Finally, the fundamental nature of these problems makes every contribution a contribution to a half a dozen fields at the same time.

2. Too Much Data

Much of our previous discussion has centered upon the difficulties encountered in the study of processes which are

incompletely described from the classical point of view. It is important then to note that in treating processes of high dimension, involving large quantities of data, complete information is as much of a handicap as a scarcity of information. As a matter of fact, the two problems are intimately related, as we shall now explain.

In using information, we must take into account the problems involved in storing it, in retrieving it when desired, in using it in an efficient fashion, and, finally, in reckoning with the cost of using the information, and the time involved. Ideally, the last two should be part of the question of efficient use.

It is clear that in many cases this means that we will not have the time to examine all the data. Consequently, we will be in the position of rejecting information, which means that we are back in the situation of incomplete information.

The general problem of recognizing, evaluating and storing information is one of the major problems in current science, and one of the major problems of our civilization. It would be important to construct various mathematical models which study the rejection of large quantities of information. A combination of the theories of mathematical statistics and differential equations may yield some useful results. We shall discuss a particular model of this type below.

3. Analogy

What we have mentioned above opens up an interesting

channel of investigation. This is the idea of proceeding by analogy. To some, use of this idea is implicit in the studies of stochastic and adaptive systems which we have mentioned above. Let us now attempt to examine the idea in more detail.

Faced with the prospect of utilizing a physical system S of such complexity that we cannot employ the actual equations describing the state of the system, or in ignorance of these equations, we construct another system S', described by fewer state variables and simpler equations, and use this system as the basis of our decisions.

It is, of course, necessary from time to time to check the conclusions based upon S' with results obtained from observation of S, and to make changes either in the conclusions or in the structure of S' on the basis of this comparison. It is well to point out that this comparison may not be as easy to effect as might be expected a priori, since the interpretation of the observation of S may itself depend upon the artificial system S' which we have constructed.

It is very possible that the human brain operates in some fashion as that described above. The problem that we wish to pose is: "How do we construct mathematical models of processes carried out in this fashion?"

4. The Techniques of Box

It is appropriate in this context to mention the interesting and effective methods proposed by Box and his colleagues, [42], [43], [44]. One of the problems that he tackles is that of

controlling a chemical engineering process where the precise chemical and physical interactions are not known.

5. The Tearing Technique of Kron

Another new and promising technique is that put forth in recent years by Kron, [45], [46], [47]. He provides systematic techniques for decomposing a complex system into a set of systems of simpler type.

6. Sequential Search

Finally, let us mention the attempts in recent years to tackle in some direct fashion the problem of determining the maximum or minimum of a function of a large number of variables. For the reasons which we have previously discussed it is clearly impossible to accomplish this in any simple enumerative way.

For the case of functions of one variable, some very elegant results exist, see [48], [49], [50]. For functions of two or more variables, optimal policies seem quite difficult to obtain, and it is reasonable to suppose that simple sub-optimal policies would be as efficient in practice. However, even these seem hard to find.

7. Routing and Switching

We have avoided all discussion of the very interesting combinatorial questions that arise in the study of optimal routing in communication and travel networks. The reader concerned with these fields may wish to refer to the expository papers [51], [52], where many additional references can be found.

REFERENCES

1. J. F. Roth, Integration in Finite Terms, Columbia Univ. Press, New York, 1948.

2. R. Bellman, Stability Theory of Differential Equations, McGraw-Hill Book Co., Inc., New York, 1954.

3. R. Bellman, A Survey of the Boundedness, Stability and Asymptotic Behavior of Linear and Nonlinear Differential and Difference Equations, Office of Naval Research, 1948.

4. V. M. Starzinskii, "Survey of works on conditions of stability of the trivial solution of a system of linear differential equations with periodic coefficients," Akad. Nauk, Prikhl. Mat. Mech., vol. 18, 1954, pp. 469-510 (translated by American Mathematical Society).

5. C. Hermite, Oeuvres, vol. 2.

6. J. Shtokalo, "On the theory of linear differential equations with quasi-periodic coefficients," Akad. Nauk. Ukraine, 1946, pp. 163-176 (Math. Rev., vol. 12, 1951, p. 334).

7. S. Lefschetz, Differential Equations: Geometric Theory, Interscience Publ., New York, 195?

8. R. Bellman, "Functional equations in the theory of dynamic programming--V: positivity and quasi-linearity," Proc. Nat. Acad. Sci. USA, vol. 41, 1955, pp. 743-746.

9. R. Kalaba, "On nonlinear differential equations, the maximum operation, and monotone convergence," J. Math. and Mech., vol. 8, 1959, pp. 519-574.

10. R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton Univ. Press, Princeton, N. J., 1960.

11. R. Bellman, M. Juncosa, and R. Kalaba, On Numerical Solution of Nonlinear Parabolic Partial Differential Equations, to appear.

12. R. E. Beckwith, Analytic and Computational Aspects of Dynamic Programming Processes, Ph.D. Thesis, Purdue Univ., 1959.

13. R. Bellman, "Some new techniques in the dynamic programming solution of variational problems," Q. Appl. Math., vol. 16, 1958, pp. 295-305.

14. R. Bellman, "On some applications of dynamic programming to matrix theory," Illinois J. Math., vol. 1, 1957, pp. 297-301.

15. R. Bellman, Dynamic Programming, Princeton Univ. Press, Princeton, N. J., 1957.

16. J. P. Roth, "An application of algebraic topology, Kron's method of tearing," Q. Appl. Math., vol. 17, 1959, pp. 1-23.

17. A. Boldyreff, Determination of the Maximal Steady State Flow of Traffic through a Railroad Network, The RAND Corporation, P-687, August 5, 1955.

18. G. Borg, "Eine umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe. Bestimmung der differential Gleichung durch die Eigenwerte," Acta Math., vol. 78, 1946, pp. 1-96.

19. I. M. Gelfand and B. M. Levitan, "On the determination of a differential equation from its spectral function," Translations Amer. Math. Soc., vol. 2, 1955, pp. 253-304.

20. R. Bellman and J. M. Danskin, A Survey of the Mathematical Theory of Time-lag, Retarded Control and Hereditary Processes, The RAND Corporation, Report R-256, 1954.

21. R. Bellman and K. Cooke, An Introduction to the Theory of Differential-difference Equations, to appear.

22. J. C. Samuels and A. C. Eringen, On Stochastic Linear Systems, Tech. Report 11, Div. Eng. Sci., Purdue Univ., 1957.

23. R. Bellman, Introduction to Matrix Analysis, McGraw-Hill Book Co., Inc., New York, 1959.

24. E. H. Kerner, "The bang structure of mixed linear lattices," Proc. Phys. Soc., A, vol. LXIX, 1956, pp. 234-244.

25. R. Bellman, "Limit theorems for non-commutative operations--I," Duke Math. J., vol. 21, 1954, pp. 491-500.

26. E. Coddington and N. Levinson, Ordinary Differential Equations, McGraw-Hill Book Co., Inc., New York, 1958.

27. L. Zadeh, "General filters for separation of signal and noise," Proc. Symp. Information Networks, Polytechnic Inst. Brooklyn, 1955, pp. 31-49.

28. N. Wiener, Nonlinear Problems in Random Theory, J. Wiley and Sons, New York, 1958.

29. R. Bellman, I. Glicksberg, and O. Gross, "On some variational problems occurring in the theory of dynamic programming," Rend. del Cir. Mate. di Palermo, ser. II, tomo III, 1954, pp. 1-35.

30. J. P. LaSalle, "On time optimal control systems," Proc. Nat. Acad. Sci. USA, vol. 45, 1959, pp. 573-577.

31. R. Bellman, I. Glicksberg, and O. Gross, "On the 'bang-bang' control problem," Q. Appl. Math., vol. 14, 1956, pp. 11-18.

32. R. Bellman, "Functional equations and successive approximation in linear and nonlinear programming," Naval Research Logistics Q., to appear.

33. R. Bellman and S. Dreyfus, "Functional approximations and dynamic programming," Math. Tables and other Aids to Comp., to appear.

34. R. Bellman, "Dynamic programming and stochastic control processes," Information and Control, vol. 1, 1958, pp. 228-239.

35. J. von Neumann and O. Morgenstern, Theory of Games and Economic Behavior, Princeton Univ. Press, Princeton, N. J., 1948.

36. J. D. Williams, The Compleat Strategyst, McGraw-Hill Book Co., Inc., New York, 1955.

37. R. Bellman, C. Clark, C. Craft, D. Malcolm, and F. Ricciardi, "On the construction of a multi-person, multi-stage business game," Operations Research, vol. 5, 1957, pp. 469-503.

38. R. Bellman, "On communication processes involving learning and random duration, 1958 IRE National Convention Record, part 4, pp. 16-20.

39. R. Bellman and R. Kalaba, "Dynamic programming and adaptive processes--I: mathematical foundation," Proc. IRE, to appear.

40. R. Bellman and R. Kalaba, "Functional equations in adaptive processes and random transmission," Trans. 1959 International Symp. on Circuit and Information Theory, pp. 271-282.

41. R. Kalman, "On a new approach to filtering and prediction problems," Amer. Soc. Mech. Eng., 1959.

42. G. E. P. Box, "Evolutionary operations, a method for increasing industrial productivity," Appl. Stat., vol. 6, 1957, pp. 3-23.

43. G. E. P. Box and J. S. Hunter, "Condensed calculations for evolutionary operation program," Technometrics, vol. 1, 1959, pp. 77-95.

44. G. E. P. Box and G. A. Coutie, "Application of digital computers in the exploration of functional relationships," Proc. Int. Elec. Eng., B-103, 1957, supplement no. 1, pp. 100-107.

45. G. Kron, "Tearing and interconnecting as a form of transformation," Q. Appl. Math., vol. 13, 1955, pp. 147-159.

46. G. Kron, "How to use the A.C. network analyzer for 'tearing,'" Matrix Tensor Q., vol. 6, 1956, pp. 131-134.

47. G. Kron, "Improved procedure for interconnecting piece-wise solutions," J. Franklin Inst., vol. 262, 1956, pp. 385-392.

48. J. Kiefer, "Optimum sequential search and approximation methods under minimum regularity assumptions," J. Soc. Ind. Appl. Math., vol. 5, 1957, pp. 105-136.

49. S. Johnson, Optimal Search is Fibonaccian, unpublished; the results are given in Dynamic Programming, pp. 34-36.

50. O. Gross and S. Johnson, "Sequential minimax search for a zero of a convex function," Math. Tables and other Aids to Comp., vol. XIII, 1959, pp. 44-51.

51. R. Bellman, "Mathematical aspects of scheduling theory," J. Soc. Ind. Appl. Math., vol. 4, 1956, pp. 168-205.

52. R. Kalaba, "On some communication network problems," Proc. Ninth Symp. Appl. Math., to appear.